

Compare Between Kriging And Fuzzy kriging (Centroid Method) With Application

Jaufar Mousa Mohammed^{1*}, Dalia Badee Omar²

¹⁻² Department of Statistics Sciences, College of Administration & Economic, University of Kirkuk, Iraq

Korespondensi penulis: jaufar@uokirkuk.edu.iq

Abstract. This research deals with a comparison of Kriging's method in predicting ordinary and fuzzy data in order to know the best method for future studies. Ordinary data was used in the field of depth of groundwater wells for 37 locations in Kirkuk city, and these data were fuzzy into fuzzy numbers of the trigonometric type that have a function of belonging, Then the centroid of each fuzzy number was found for the purpose of facilitating the calculations. An unknown point was predicted for both the ordinary and fuzzy data. After comparing the results with a standard of least variance, it was found that the fuzzy data had better results than the ordinary data.

Keywords: Kriging, Fuzzy kriging, Centroid method.

1. INTRODUCTION

The Kriging method is one of the advanced methods in spatial statistics in predicting unknown points. Each variable deals with the spatial phenomenon on the basis of location and distance. These phenomena have fixed, definite and well-known measurements and can be controlled resulting errors which are random variables by studying their behavior. Recently the theory of fuzzy aggregates has appeared, which is concerned with phenomena whose variables cannot be measured by points but are measured in the form of periods, or what is described as uncertain cases or cases with fuzzy data because of their characteristics that make them unclear, such as variables that belong in certain proportions to their groups and do not have full affiliation, as well as linguistic variables that cannot be measured numerically and there are variables that are measured approximately but they are in fact fuzzy. Such measurements in spatial applications require special formulation because they are used to measurements that are unclear. On this basis the idea of studying came how to formulate the fuzzy kriging method for fuzzy or unclear data ,based on fuzzy logic information and there are several methods for this idea, one of these methods that was used in this research is the method of finding the centroid value for the fuzzy number that is converting the fuzzy number in to real number ,hence use the Fuzzy Kriging method for estimation and compare it with the ordinary data and compare the results between the two cases (ordinary and fuzzy), using the least variance .

2. BASIC CONCEPTS OF FUZZY LOGIC

A. Crisp Set

A set whose elements have a constant property, which takes one of two values (1) when the element belongs to the set and (0) when the element does not belong to the set.

Let A be a set known as a function and it called the characteristic function as follows as: (Ross, 2005)

$$\chi_A(x) = \begin{cases} 0 & \text{if } x \notin A \\ 1 & \text{if } x \in A \end{cases}$$

B. Fuzzy Set

It is the group whose elements have a degree of affiliation or a membership degree and are real numbers within the closed interval $[0,1]$. The degree of membership is expresses by the affiliation function or the membership function, which represents the degree of belonging of the variable x to the fuzzy group and is written in the following form:

$$A = \{(x, \mu_A(x)) : x \in X : \mu_A(x) : X \rightarrow [0, 1]\}$$

If it is in an intermittently state, it will be as follows:

$$A = \sum_{i=1}^n \frac{\mu_A(x_i)}{x_i} = \frac{\mu_A(x_1)}{x_1} + \frac{\mu_A(x_2)}{x_2} + \dots + \frac{\mu_A(x_n)}{x_n}$$

If it a continuous state, it is written as follows:

$$A = \left\{ \int \frac{\mu_A(x)}{x} \right\}$$

(α -Cut Set)

Let A be a fuzzy subset of the universal set X , then we define (α -cut set) on A as follows.

$$A_{[\alpha]} = \{x \in X : \mu_A \geq \alpha\} , \quad \alpha \in [0, 1]$$

If it is as follows:

$$A_{[\alpha^+]} = \{x \in X : \mu_A > \alpha\} , \quad \alpha \in [0, 1]$$

This is called the α -cut set Strong on A .

C. Support of Fuzzy Set

The predicate $\text{Sup}(A)$ of fuzzy set A is defined as the set of elements in the sup set X , that have membership function with non-zero value, meaning that:

$$\text{Sup}(A) = \{x \in X: \mu_A(x) > 0\}$$

D. Hight of Fuzzy Set

The height of the fuzzy group A is defined as the highest membership degree of any element of the fuzzy group A and denoted by $h(A)$, meaning that

$$h(A) = \max\{\mu_A(x); \text{for any } x \in A\}$$

E. The Core of Fuzzy Set

A is the fuzzy subset of the general set X as normal if:

$$\text{Core}(A) = \{x \in A: \mu_A(x) = 1\}$$

Normal Fuzzy Set

The fuzzy set is normal if the highest value of the membership function is equal to 1 or $h(A)=1$, vice versa the fuzzy set is called sub normal.

Convex Fuzzy Set

The fuzzy subset A of the universal set X is called convex if:

$$t = \lambda r + (1 - \lambda)s, \mu_A(t) \geq \min[\mu_A(r), \mu_A(s)] \text{ where } r, s \in R, \lambda \in [0, 1] \cup [9].$$

Fuzzy number

Fuzzy number A is fuzzy subset in the universal set X if it :

1. A is Normalized fuzzy set.
2. A is convex fuzzy set.
3. It has an affiliation function and is semi-continuous
4. Knowledge of real numbers R . (Chen and Pham 2000)

Triangular Fuzzy Number

The fuzzy subset A in the universal set X is called the trigonometric fuzzy number which is expressed in the form $A=(a,b,c)$ where $a < b < c$ if it has an affiliation function as follows:

$$\mu_A(x) = \begin{cases} \frac{x-a}{b-a} & \text{if } a \leq x \leq b \\ \frac{c-x}{c-b} & \text{if } b \leq x \leq c \\ 0 & \text{other wise} \end{cases} \dots\dots\dots (1)$$

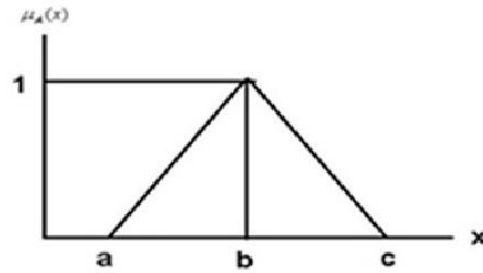


Figure (1) Triangular Fuzzy Number

(A) can be converted to a crisp number by using the centroid method, this process is called (defuzzification) and the centroid method is written in the following form:

$$A_c = \frac{\int x \mu_A(x) dx}{\int \mu_A(x) dx} = \frac{1}{3} (a + b + c) \dots \dots \dots (2)$$

This method designed by M.Sugeno in 1985 is the most widely used and it is a very accurate technique.[5][8]

1. Spatial Variable

The variables that spatial statistics deals with are different from the normal variables, as each value of the locational variable has coordinates that represent the location of that point whether it is on the surface of the earth in the plane or in the ground, or outside the earth for example (air pollution with gases). Suppose that $z(x)$ represents the spatial variable at the location x within the region D in the conventional space $x \subseteq D \subseteq R^p$ where x is in the plane $p = 2$ or $p = 3$ in space. This variable can be measured on a sample of size n from locations, and these measurements are symbolized by the variable $z(x)$ and its values are $z(x_1), z(x_2), z(x_3), \dots, z(x_n)$ in locations $x_1, x_2, x_3, \dots, x_n$ respectively. These variables are separated by the displacement h and the distance between the variables can be calculated using the traditional distance method. [2]

a) Stationary second order

The random variable $z(x)$ is called stationary second order if the expectation exists and does not depend on the location x , that is:

$$E[z(x)] = \mu, \quad \forall x \in D$$

And that a pair of spatial random variables $[z(x), z(x+h)]$ the covariance is defined and depends on the dispersion h only, that is:

$$C(h) = E[z(x) - z(x+h) - \mu^2] = Cov[z(x), z(x+h)], \quad \forall x, x+h \in D$$

b) Semivariogram function

It is the average square of the differences between spatial observations that are separated by a displacement h , where $n(h)$ represents the number of pairs of observations that $z(x_i)$, $z(x_{i+h})$ are separated by a displacement of h , which is of the form:

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} ((z(x_i) - z(x_i + h))^2) \dots \dots \dots (3)$$

The reason for studying the variogram function is that its mathematical formula represents the variance of differences between spatial observations that are far from each other by displacement h . when greater the displacement h between the observations, the greater the covariance becomes large until its height stabilizes at a certain distance such as $h = a$, and this distance a is called the range, and after that we notice the disappearance of the covariance in the variogram function, where it stabilizes with a constant value equal to the variance σ^2 of the observations and this variance is called (Sill) as shown in figure (2).

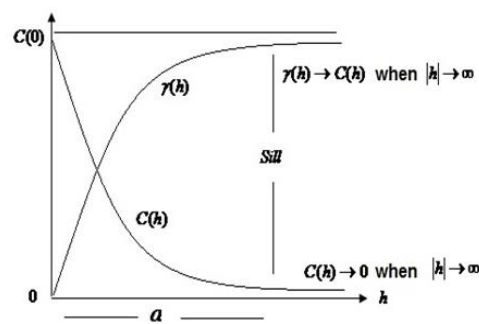


Figure (2) Variogram And Covariance

When h near zero from the right side, the semi-variogram function is not equal to zero, but has a value equal to (ψ_0) and this phenomenon represents the lack of continuity or discontinuity of the variogram function at $h=0$. It is called in the field of spatial statistic Nugget Effect it represents random errors in unite of measurement when the displacement changes suddenly from units of millimeters to units of meters or kilometers, figure (3). [1], [3]

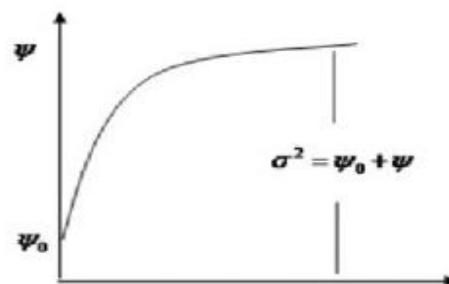


Figure (3) Negget Effect

There is a relationship between the covariance function $C(h)$ and the variogram function $\gamma(h)$ variance σ^2 , where :

$C(0) = \sigma^2$ which represents the variance.

There are multiple versions of the variogram function, including the spherical variogram

$$\gamma(h) = \begin{cases} \psi_0 & h = 0 \\ \psi_0 + \psi \left[\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right] & 0 < h \leq a \\ \psi_0 + \psi & h > a \end{cases} \dots \dots \dots (4)$$

And from the relationship:

$$\gamma(h) = C(0) - C(h)$$

We obtain the spherical model in terms of the covariance function in the following form:

$$C(h) = \begin{cases} \psi_0 + \psi & h = 0 \\ \psi \left[1 - \frac{3h}{2a} + \frac{1}{2} \left(\frac{h}{a} \right)^3 \right] & 0 < h \leq a \\ 0 & h > a \end{cases} \dots \dots \dots (5)$$

Whereas ψ , a requires estimation [10]

2. Ordinary kriging approach and terminology

The basic linear kriging estimator for $z(x_0)$ is defined in the following form:

$$\hat{z}(x_0) - m(x) = \sum_{i=1}^{n(x)} \lambda_i [z(x_i) - m(x_i)]$$

Where

x_0, x_i : location of the estimated point and one of its adjacent data points.

$n(x)$: the number of adjacent data points used to estimate $\hat{z}(x_0)$.

$m(x_0), m(x_i)$: are the expectation values (mean) for $z(x_0)$ and $z(x_i)$ respectively, and $\lambda_i(x)$ represents the kriging weights for the original $z(x_i)$ values used to estimate the location x_0 and these weights for the original values that vary location x_0 to be estimated.

$z(x)$: is a random field composed of the direction $m(x)$ and the residuals $R(x)$ that is $z(x) = m(x) + R(x)$, and kriging estimates residuals for x is the sum of the residual weights of other adjacent data points. Kriging's weights $\lambda_i(x)$ are derived from the covariance function or from the semi - variogram function which should describe the components of the residuals.

The objective of the above equation is to determine the weights $\lambda_i(x)$ that make the variance of the kriging estimator as small as possible.

$$\sigma_E^2 = \text{var}[\hat{z}(x_0) - z(x_0)] \quad \min \dots \dots \dots (6)$$

Under unbiased conditions : $E[\hat{z}(x_0) - z(x_0)] = 0$

3. Ordinary kriging

In this method, we assume that the mean is unknown in the study area D , but is constant in the partial $A \subset D$ region, that is, in the location of the points adjacent to the point to be estimated. This means that $\forall x \in A \subset D, m(x_i) = m(x_0)$ for the values of the points nearby data. In this case, $z(x_i)$ is used to estimate $z(x_0)$ that is not measured. Thus, ordinary kriging estimator is in the following form:

$$\begin{aligned}\hat{z}(x_0) &= m(x) + \sum_{i=1}^{n(x)} \lambda_i [z(x_i) - m(x_i)] \\ &= \sum_{i=1}^{n(x)} \lambda_i z(x_i) + \left[1 - \sum_{i=1}^{n(x)} \lambda_i\right] m(x_0)\end{aligned}$$

In order for this estimator to be unbiased, the sum of the weights of the ordinary kriging λ_i must be equal to one, so the estimator of ordinary kriging unbiased as follows:

$$\hat{z}(x_0) = \sum_{i=1}^{n(x)} \lambda_i z(x_i) \quad \text{with} \quad \sum_{i=1}^{n(x)} \lambda_i = 1 \dots \dots (7)$$

In order to reduce the variance error:

$$\sigma_E^2 = \text{var}[\hat{z}(x_0) - z(x_0)] = \text{var}[z(x_0)] + \text{var}[\hat{z}(x_0)] - 2\text{Cov}[z(x_0), \hat{z}(x_0)]$$

Under the constraint $\sum_{i=1}^{n(x)} \lambda_i = 1$, we add the Lagrange factorial to the above amount as follows:

$$L = \sigma_E^2 + 2\mu \left[1 - \sum_{i=1}^{n(x)} \lambda_i\right]$$

And by taking the partial derivative with respect to λ , μ and set it equal to zero, we get:

$$\frac{1}{2} \frac{\partial L}{\partial \lambda} = \sum_{i=1}^{n(x)} \lambda_i C(z(x_j) - z(x_i)) + \mu - C(z(x_i) - z(x_0)) = 0$$

$$\frac{1}{2} \frac{\partial L}{\partial \mu} = 1 - \sum_{i=1}^{n(x)} \lambda_i = 0$$

Thus, the system of Ordinary Kriging equations is in the following form:

$$\begin{cases} \sum_{i=1}^{n(x)} \lambda_i C(z(x_i) - z(x_j)) + \mu = C(z(x_i) - z(x_0)) & j = 1, 2, \dots, n(x) \\ \sum_{i=1}^{n(x)} \lambda_i = 1 \end{cases}$$

Which can be written as follows:

$$C\lambda = B$$

$$\lambda = C^{-1}B \dots \dots \dots (8)$$

When (B) is the vector of the covariance function, it is found between the point to be estimated with the points adjacent to it, after finding the inverse of the square matrix C we can know the Lagrange multiplier μ and the weights λ that are used in the linear structure of Kriging in the prediction and the variance of the Kriging estimator is as follows:

$$\sigma_{ok}^2 = C(0) - \sum_{i=1}^{n(x)} \lambda_i C(z(x_i) - z(x_0)) - \mu \dots \dots \dots (9)$$

[11], [7]

4. Analysis fuzzy spatial data

In this part of the paper, we present a combined work of analyzing fuzzy spatial data: the first is to remove fuzzy from the fuzzy data and the second is to use Kriging prediction, which is used to sample fuzzy spatial data. Although many of the available data are not of accurate value, However, it can be modeled through the fuzzy sets and dealt with in a more effective method. [4]

5. Fuzzy spatial variable

If we assume that the spatial data available to us are fuzzy numbers of triangular type and have a function of belonging, as in formula (1), then dealing with such numbers is complicated, so there are methods in fuzzy logic that can be used to facilitate the process of dealing with them, such as finding the centroid of the fuzzy numbers of the data so that it becomes in the form $z_c(x_1), z_c(x_2), \dots, z_c(x_n)$ at stationary locations $(x_1, x_2, \dots, x_n)$ such data is called fuzzy data with crisp values, which means that has been removed the fuzzy using formula (2). [6]

6. Fuzzy ordinary kriging by centroid method

After calculating the centroid of the fuzzy data and converting it into crisp values, Kriging estimator in (7) can be formulated as follows:

$$\hat{z}_C(x_0) = \sum_{i=1}^{n(x)} \lambda_i z_C(x_i) \quad \text{with} \quad \sum_{i=1}^{n(x)} \lambda_i = 1 \dots \dots \dots (10)$$

The weights $\lambda_i(x)$ are determined to make the variance of the fuzzy Kriging estimator as less as possible:

$$\sigma_E^2 = var[\hat{z}_C(x_0) - z_C(x_0)] \quad \min \dots \dots \dots (11)$$

As in the previous paragraph, the system equations of kriging fuzzy ordinary is in the following form:

$$\begin{cases} \sum_{i=1}^{n(x)} \lambda_i C(z_C(x_i) - z_C(x_j)) + \mu = C(z_C(x_i) - z_C(x_0)) , j = 1, 2, \dots, n(x) \\ \sum_{i=1}^{n(x)} \lambda_i = 1 \end{cases}$$

which can be written as follows

$$C\lambda = B$$

$$\lambda = C^{-1}B \dots \dots \dots (12)$$

For fuzzy data

When we get the fuzzy kriging weights and the Lagrange multiplier parameter, the variance of the fuzzy kriging estimator is as follows:

$$\sigma_{ok}^2 = C(0) - \sum_{i=1}^{n(x)} \lambda_i C(z_C(x_i) - z_C(x_0)) - \mu \dots \dots \dots (13)$$

[12], [6]

الجانِب التطبيقِي

“Application Part”

Due to the importance of prediction in spatial statistics in many practical fields, which takes into account the basic spatial structure found in geological information (distances between locations), and by means of which it is possible to statistically infer optimal predictions for certain unmeasured locations, so this research was applied to real data in the field of water depth Wells for 37 locations in Kirkuk city and this data has been obtained from the Groundwater Authority in the governorate affiliated to the Ministry of Water Resources and shown in Table (1) in the appendix, where we note that each location in this data contains the coordinates of $u(x)$, which represents the east West and $v(x)$ which represents north-south and the values of $z(x_i)$ which represents the amount of depth of the wells and in order to compare between the normal data and the fuzzy data to know which type of data gives better results in prediction, the data in Table (1) has been fuzzed to the fuzzy numbers trigonometric has a affiliation function as in formula (1), figure (1) using the method of finding the centroid mentioned in formula (2), the fuzzy data was converted into crisp values in order to facilitate the calculations.

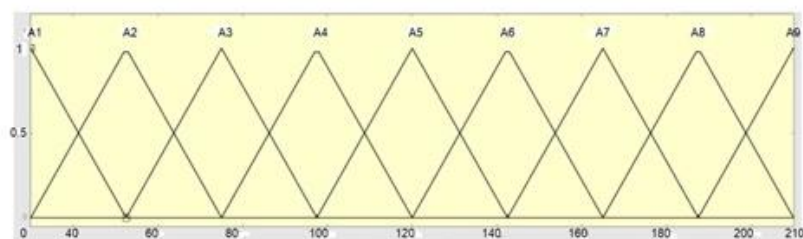


Figure (4) Fuzzy Number

جدول (1) البيانات الاعتيادية والضبابية

Location	$u(x)$	$v(x)$	Depth $z(x)$ (Crisp Data)	Depth $z_c(x)$ (Fuzzy Data)
1	35.46	44.38	150	142.5
2	35.45	44.38	190	187.5
3	35.46	44.4	80	75
4	35.44	44.37	102	97.55
5	35.44	44.4	98	97.55
6	35.44	44.41	95	97.55
7	35.41	44.39	108	97.55
8	35.43	44.33	95	97.55
9	35.41	44.41	135	142.5
10	35.45	44.32	90	97.55
11	35.4	44.4	102	97.55
12	35.51	44.34	180	187.5
13	35.38	44.35	101	97.55
14	35.38	44.35	96	97.55
15	35.39	44.29	201	210
16	35.6	44.34	75	75
17	35.3	44.37	113	120
18	35.63	44.4	88	97.55
19	35.3	44.51	102	97.55
20	35.58	44.17	118	120
21	35.46	44.13	95	97.55
22	35.43	44.13	82	75
23	35.49	44.63	125	120
24	35.59	44.13	115	120
25	35.73	44.47	115	120
26	35.67	44.01	137	142.5
27	35.38	43.94	58	52.51
28	35.3	44.8	40	30
29	35.71	43.89	130	120
30	35.44	43.78	84	75
31	34.96	43.9	190	187.5
32	35.64	45.13	107	97.55
33	35.02	45.12	120	120
34	34.58	44.38	47	52.51
35	35.41	45.38	116	120
36	35.09	45.37	84	75
37	33.13	44.45	80	75

The point in location (1) whose coordinates are (35.46, 44.38) both normal and fuzzy data was studied for the purpose of predicting them, where the variogram function was calculated for the normal and fuzzy data and we obtained the results shown in Tables (3) and (4) and then a graph was found between The **variogram** function $\gamma(h)$ and the displacement (h) in the figure below:

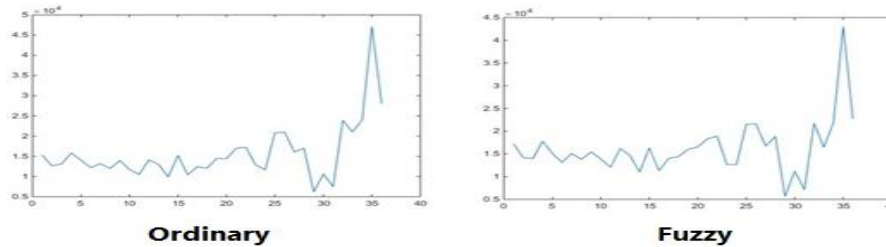


Figure (5) Simvariogram Function

From the above figure it is shown that the best mathematical model for the normal and fuzzy case of this study is the spherical model as in formula (5) and from the drawing also the parameters ψ , ψ_0 , were estimated as in the table below:

جدول (2) نتائج تقدير المعلمات

ψ	ψ_0	a	نوع البيانات
2.8	1.5	0.071	الاعتيادي
2.6	1.7	0.071	الضبابي

Thus the formula for the covariance function for the normal case is as follows:

Normal case:

$$C(h) = \begin{cases} 4.3 & h = 0 \\ 2.8 \left[1 - \frac{3h}{2 * 0.071} + \frac{1}{2} \left(\frac{h}{0.071} \right)^3 \right] & 0 < h \leq 0.071 \\ 0 & h > 0.071 \end{cases}$$

Fuzzy case:

$$C(h) = \begin{cases} 4.3 & h = 0 \\ 2.6 \left[1 - \frac{3h}{2 * 0.071} + \frac{1}{2} \left(\frac{h}{0.071} \right)^3 \right] & 0 < h \leq 0.071 \\ 0 & h > 0.071 \end{cases}$$

From these two models, the covariance matrix was obtained between the normal and fuzzy data points, respectively, which must be positive, and then a vector of covariance was found between the point to be estimated with the data points, and thus we got the weights λ_i and by using the linear structure of Kriging, the prediction is made for the unmeasured point

that has coordinates (35.46, 44.38), we note that the prediction value for the normal data $\hat{z}(x_0)$ =128.8605 and the fuzzy $\hat{z}_c(x_0)$ =127.33, We also note that the value of the estimated variance σ_{ok}^2 in the fuzzy case is σ_{ok}^2 =2.19 which is less than normal, which is equal to σ_{ok}^2 = 2.3073 and this indicates a preference for the fuzzy data in this study, Tables (5) and (6).

3. CONCLUSIONS AND RECOMMENDATIONS

- 1- Normal spatial data can be blurred into trigonometric spatial fuzzy data.
- 2- Using the Kriging method to estimate the fuzzy spatial data after finding centroid value.
- 3- After comparing the results of the estimation using the Kriging method between the fuzzy spatial data of the trigonometric type and the normal spatial data, it was found that there is a preference for the fuzzy spatial data.
- 4- The sum of the weights is $\sum_{i=1}^n \lambda_i = 1$ for both the normal and fuzzy spatial data.
- 5- We recommend using the Kriging method to estimate fuzzy Caussian and trapezoidal spatial data and compare the results with normal spatial data.
- 6- It is preferable that the sample size be large for the data used to estimate, as this gives better results.
- 7- Choose the appropriate model through the graph between the variogram function and the displacement h , otherwise we will get inaccurate results.

REFERENCES

- 1- Ali Mohammed Ali Chichan, Mustafa Ali Fakhri, Zana najim(2021). "**Comparison Between Iterative Maximum Likelihood Estimators Method And Bayesian Method For Estimating Logistic Regression Model Parameters With Practical Application** ".journal of Economics And Administrative Sciences, Published by Kirkuk University,2015 volume(10)Issue (2), pp(259-270).
- 2- Bardossy, A., Bogardi, I., & Kelly, W. E. (1990). Kriging with imprecise (fuzzy) variograms. I: Theory. *Mathematical Geology*, 22, 63-79.
- 3- Bohling, G. (2005). Introduction to geostatistics and variogram analysis. *Kansas geological survey*, 1, 1-20.
- 4- Guo, D., Guo, R., & Thiart, C. (2003). Integrating GIS with fuzzy logic and geostatistics: predicting air pollutant pm10 for California, using fuzzy kriging. *Doctoral dissertation, MSc. Thesis*.

- 5- Hesamian, G., & Akbari, M. G. (2020). A kriging method for fuzzy spatial data. *International Journal of Systems Science*, 51(11), 1945-1958.
- 6- Lee, K. H. (2004). "**First course on fuzzy theory and applications**", (Vol. 27). Springer Science & Business Media.
- 7- Mohammed, J. M., Tawfeeq, A. F., & Aziz, M. M. (2022). Maximum likelihood for fuzzy pure spatial autoregressive model. *International Journal of Nonlinear Analysis and Applications*, 13(1), 2267-2277.
- 8- Nielsen, A. A. (2009). Geostatistics and analysis of spatial data. *Informatics and Mathematical Modelling, Technical University of Denmark, DTU*, 7-12.
- 9- Quevedo, J. R. N. (2017). "**Fuzzy sets. A way to represent ambiguity and subjectivity**" . Boletín de Matemáticas, 24(1), 57-89.
- 10- Ross, T. J. (2005) . "**Fuzzy logic with engineering application**" . John Wiley & Sons
- 11- Studzinski, J. (2011). Application of Kriging Algorithms for Solving some Water Nets Management Tasks. In *EnviroInfo* (pp. 477-488).
- 12- Tony E. Smith (2015): "**Spatial Data Analysis** " University Pennsylvania School of engineering and applied science. tesmith@seas.upenn.edu.
- 13- Zhang, A. (2016). Statistical analysis of fuzzy linear regression model based on centroid method. *Applied Mathematics*, 7(7), 579-586